S&H Form: FORM PTO-1390 (2/01)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER
1454.1218

10/069989

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/DE00/01787	31 May 2000	3 September 1999

TITLE OF INVENTION
METHOD FOR DETECTING AND EVALUATING WORD SPEECH SIGNALS REPRESENTING A WORD FROM A USER OF A SPEECH RECOGNITION SYSTEM

APPLICANT(S) FOR DO/EO/US
Josef BAUER et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.
2. ☒ This is an express request to immediately begin national examination procedures (35 U.S.C. 371(f)).
3. ☒ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
4. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☒ is transmitted herewith (required only if not transmitted by the International Bureau).
   b. ☐ has been transmitted by the International Bureau.
   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
5. ☒ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
6. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
   b. ☐ have been transmitted by the International Bureau.
   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US)
7. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
8. ☒ An oath or declaration of the inventor (35 U.S.C. 371(c)(4)).
9. ☒ A translation of the Annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 10-15 below concern document(s) or information included:

10. ☒ An Information Disclosure Statement Under 37 CFR 1.97 and 1.98.
11. ☒ An assignment document for recording.
    Please mail the recorded assignment document to:
    a. ☒ the person whose signature, name & address appears at the bottom of this document.
    b. ☐ the following:
12. ☒ A preliminary amendment.
13. ☒ A substitute specification
14. ☐ A change of power of attorney and/or address letter.
15. ☒ Other items or information:

Copies of PCT/EASY forms filed with International Application, cover page of International Application as published, International Search Report, and International Preliminary Examination Report.

☒ The U.S. National Fee (35 U.S.C. 371(c)(1)) and other fees as follows:

CLAIMS	(1) FOR	(2) NUMBER FILED	(3) NUMBER EXTRA	(4) RATE	(5) CALCULATIONS
	TOTAL CLAIMS	41 -20=	21	x $ 18.00	378.00
	INDEPENDENT CLAIMS	12 -3=	9	x $ 84.00	756.00
	MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+$280.00	0.00
	BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(4): ☐ Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO .....$1,040 ☒ International preliminary examination fee (37 C.F.R. 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO.....$ 890 ☐ International preliminary examination fee (37 C.F.R. 1.482) not paid to USPTO but international search fee (37 C.F.R. 1.445(a)(2) paid to USPTO...$ 740 ☐ International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provision of PCT Article 33(1)-(4).....$ 710 ☐ International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2) to (4) .....$ 100				890.00
	Surcharge of $130 for furnishing the National fee or oath or declaration later than ☐ 20 ☐ 30 mos. from the earliest claimed priority date (37 CFR 1.482(e)).				0.00
			TOTAL OF ABOVE CALCULATIONS		2,024.00
	Reduction by 1/2 for filing by small entity, if applicable. Affidavit must be filed also. (Note 37 CFR 1.9, 1.27, 1.28.)				
			SUBTOTAL		2,024.00
	Processing fee of $130 for furnishing the English Translation later than [ ] 20 [ ] 30 mos. from the earliest claimed priority date (37 CFR 1.482(f)).				
			TOTAL NATIONAL FEE		2,024.00
	Fee for recording the enclosed assignment (37 CFR 1.21(h)).				+ 40.00
			TOTAL FEES ENCLOSED		2,064.00

a. ☒ A check in the amount of $2,064.00 to cover the above fees is enclosed.
b. ☐ Please charge my Deposit Account No. 19-3935 in the Amount of $ to cover the above fees. A duplicate copy of this sheet is enclosed.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 19-3935. A duplicate copy of this sheet is enclosed.

21171
PATENT TRADEMARK OFFICE

SUBMITTED BY: STAAS & HALSEY LLP

Type Name	Richard A. Gollhofer	Reg. No.	31,106
Signature	Richard A. Gollhofer	Date	3/1/02

Docket No.: 1454.1218

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

Josef BAUER et al.

Serial No. Group Art Unit:

Confirmation No.

Filed: (concurrently) Examiner:

For: SYSTEM FOR DETECTING AND EVALUATING WORD SPEECH SIGNALS REPRESENTING A WORD FROM A USER OF A SPEECH RECOGNITION SYSTEM (as amended)

PRELIMINARY AMENDMENT

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Before examination of the above-identified application, please amend the application as follows:

IN THE TITLE:

Please DELETE the Title in its entirety and REPLACE with the following new Title.

-- SYSTEM FOR DETECTING AND EVALUATING WORD SPEECH SIGNALS REPRESENTING A WORD FROM A USER OF A SPEECH RECOGNITION SYSTEM --.

IN THE SPECIFICATION:

Please REPLACE the pending specification with the substitute specification attached hereto.

IN THE CLAIMS:

Please cancel without prejudice or disclaimer claims 1-13 in the underlying PCT application and ADD new claims 14-54 in accordance with the following:

14. (NEW) A method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

detecting acoustic word speech signals from a user;

carrying out a speech recognition operation using a first vocabulary;

assessing probability of correct speech recognition;

prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

detecting and evaluating letter signals as input by the user;

carrying out a word recognition operation, after said detecting and evaluating of respective letter signals representing a single letter, using a second vocabulary larger than the first vocabulary;

assessing the probability of correct word recognition; and

terminating spelling and outputting a word obtained with a second desired probability by said assessing the probability of correct word recognition.

15. (NEW) The method as claimed in claim 14, wherein the word recognition operation includes

assigning a letter recognition probability based on the letter speech signals; and

determining a word list of all words in the second vocabulary having a letter recognition probability not lower than a highest determined letter recognition probability for any word, minus a first threshold value.

16. (NEW) The method as claimed in claim 15,

wherein said assessing the probability of correct word recognition comprises determining whether the word list contains only a single word, and

wherein said terminating spelling and outputting the word is performed if only a single word is contained in the word list.

17. (NEW) The method as claimed in claim 16,

further comprising:

carrying out speech recognition of the word speech signals using the word list with each word assigned a speech recognition probability; and

determining whether a highest speech recognition probability and a second highest speech recognition probability differ from one another by a predetermined threshold value; and

wherein if the predetermined threshold value is exceeded by a difference between the highest and second highest speech recognition probabilities, said terminating

spelling and outputting the word is performed for the word in the word list with the highest speech recognition probability.

18. (NEW) A method for detecting and evaluating word speech signals representing a word from the user of a speech recognition system, comprising:

detecting acoustic word speech signals from a user;

carrying out a speech recognition operation;

assessing probability of correct speech recognition;

prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

detecting and evaluating letter signals as input by the user;

carrying out a word recognition operation, after said detecting and evaluating of respective letter signals representing a single letter;

assessing the probability of correct word recognition;

terminating spelling and outputting a word obtained with a second desired probability by said assessing the probability of correct word recognition; and

carrying out speech recognition of the word speech signals using the letter signals as detected and evaluated, if the correct word recognition is not obtained with the second desired probability.

19. (NEW) A method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

detecting acoustic word speech signals from a user;

carrying out a speech recognition operation to obtain a speech recognition probability;

assessing probability of correct speech recognition;

prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

detecting and evaluating letter signals as input by the user of at least one letter, to obtain a letter recognition probability based on each detected letter signal;

carrying out a word recognition operation, after said detecting and evaluating of respective letter signals representing a single letter, based on a combined recognition probability using the letter recognition probability and the speech recognition probability;

assessing the probability of correct word recognition; and

terminating spelling and outputting a word if the word is obtained with a second desired probability by said assessing the probability of correct word recognition.

20. (NEW) The method as claimed in claim 19, further comprising generating a word list based on the combined recognition probability.

21. (NEW) The method as claimed in claim 20, wherein said terminating spelling and outputting the word is based solely on a single interrogation as to whether the combined recognition probability is the second desired recognition probability.

22. (NEW) The method as claimed in claim 21, wherein said terminating spelling and outputting the word includes

outputting an appropriate message to the user; and

terminating said detecting of the acoustic word speech signals.

23. (NEW) The method as claimed in claim 22, further comprising, after said detecting and evaluating of the letter speech signals respectively representing a letter:

determining whether the user is continuing to speak;

continuing said detecting and evaluating and the word recognition operation for next speech signals respectively representing a letter, if the user continues to speak; and

outputting one of the word list and a predetermined number of the words with highest probabilities in the word list, if the user does not continue to speak.

24. (NEW) A device for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

speech detection means for detecting acoustic word speech signals from a user;

initial speech recognition means for carrying out a speech recognition operation using a first vocabulary;

speech assessment means for assessing probability of correct speech recognition;

means for prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

letter detection means for detecting and evaluating letter signals as input by the user;

word recognition means for carrying out a word recognition operation, after said letter detection means has evaluated respective letter signals representing a single letter, using a second vocabulary larger than the first vocabulary;

word assessment means for assessing the probability of correct word recognition; and

termination means for terminating spelling and outputting a word obtained with a second desired probability by said word assessment means.

25. (NEW) The device as claimed in claim 24, wherein said word recognition means includes

means for assigning a letter recognition probability based on the letter speech signals; and

means for determining a word list of all words in the second vocabulary having a letter recognition probability not lower than a highest determined letter recognition probability for any word, minus a first threshold value.

26. (NEW) The device as claimed in claim 25,

wherein said word assessment means comprises means for determining whether the word list contains only a single word, and

wherein said termination means terminates spelling and outputs the word if only a single word is contained in the word list.

27. (NEW) The device as claimed in claim 26,

further comprising:

supplemental speech recognition means for carrying out speech recognition of the word speech signals using the word list with each word assigned a speech recognition probability; and

means for determining whether a highest speech recognition probability and a second highest speech recognition probability differ from one another by a predetermined threshold value; and

wherein if the predetermined threshold value is exceeded by a difference between the highest and second highest speech recognition probabilities, said termination means terminates spelling and outputs the word in the word list with the highest speech recognition probability.

28. (NEW) A device for detecting and evaluating word speech signals representing a word from the user of a speech recognition system, comprising:

speech detection means for detecting acoustic word speech signals from a user;

initial speech recognition means for carrying out a speech recognition operation;

speech assessment means for assessing probability of correct speech recognition;

means for prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

letter detection means for detecting and evaluating letter signals as input by the user;

word recognition means for carrying out a word recognition operation, after said letter detection means has evaluated respective letter signals representing a single letter;

word assessment means for assessing the probability of correct word recognition;

termination means for terminating spelling and outputting a word obtained with a second desired probability by said word assessment means; and

supplemental speech means for carrying out speech recognition of the word speech signals using the letter signals as detected and evaluated, if the correct word recognition is not obtained with the second desired probability.

29. (NEW) A device for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

speech detection means for detecting acoustic word speech signals from a user;

speech recognition means for carrying out a speech recognition operation to obtain a speech recognition probability;

speech assessment means for assessing probability of correct speech recognition;

means for prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

letter detection means for detecting and evaluating letter signals as input by the user of at least one letter, to obtain a letter recognition probability based on each detected letter signal;

word recognition means for carrying out a word recognition operation, after said letter detection means has evaluated respective letter signals representing a single letter, based on a combined recognition probability using the letter recognition probability and the speech recognition probability;

word assessment means for assessing the probability of correct word recognition; and

termination means for terminating spelling and outputting a word if the word is obtained with a second desired probability by said word assessment means.

30. (NEW) The device as claimed in claim 29, further comprising means for generating a word list based on the combined recognition probability.

31. (NEW) The device as claimed in claim 30, wherein said termination means bases termination of spelling and outputting of the word on a single interrogation as to whether the combined recognition probability is the second desired recognition probability.

32. (NEW) The device as claimed in claim 31, wherein said termination means comprises:

means for outputting an appropriate message to the user; and

means for terminating detection of the acoustic word speech signals.

33. (NEW) The device as claimed in claim 32, further comprising:

means for determining whether the user is continuing to speak;

means for continuing the detection, the evaluation and the word recognition operation for next speech signals respectively representing a letter, if the user continues to speak; and

means for outputting one of the word list and a predetermined number of the words with highest probabilities in the word list, if the user does not continue to speak.

34. (NEW) A communication device for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

a data bus;

at least one memory device, coupled to said data bus, to store at least one vocabulary and at least one program;

a speech recognition processor, coupled to said data bus, to detect acoustic word speech signals from a user and to carry out a speech recognition operation using a first vocabulary;

a speech output device, coupled to said data bus, to produce audio signals simulating speech; and

a central processor, coupled to said data bus, to assess probability of correct speech recognition, to generate first output signals causing said speech output device to prompt the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability, to evaluate letter signals as input by the user, to carry out a word recognition operation following evaluation of respective letter signals representing a single letter using a second vocabulary larger than the first vocabulary, to assess the probability of correct word recognition, and to terminate the evaluation of respective letter signals and generate second output signals causing said speech output device to output a word obtained with a second desired probability based upon assessment of the probability of correct word recognition.

35. (NEW) The communication device as claimed in claim 34, wherein said central processor also assigns a letter recognition probability based on the letter speech signals and determines a word list of all words in the second vocabulary having a letter recognition probability not lower than a highest determined letter recognition probability for any word, minus a first threshold value.

36. (NEW) The communication device as claimed in claim 35, wherein said central processor assesses the probability of correct word recognition by determining whether the word list contains only a single word, and terminates spelling and causes output of the word if only a single word is contained in the word list.

37. (NEW) The communication device as claimed in claim 36, wherein said central processor also carries out speech recognition of the word speech signals using the word list with each word assigned a speech recognition probability and determines whether a highest speech recognition probability and a second highest speech recognition probability differ from one another by a predetermined threshold value, and, if the predetermined threshold value is exceeded by a difference between the highest and second highest speech recognition

probabilities, said central processor terminates spelling and causes output of the word in the word list with the highest speech recognition probability.

38. (NEW) A communication device for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

a data bus;

at least one memory device, coupled to said data bus, to store at least one program;

a speech recognition processor, coupled to said data bus, to detect acoustic word speech signals from a user and to carry out a speech recognition operation;

a speech output device, coupled to said data bus, to produce audio signals simulating speech; and

a central processor, coupled to said data bus, to assess probability of correct speech recognition, to generate first output signals causing said speech output device to prompt the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability, to evaluate letter signals as input by the user, to carry out a word recognition operation following evaluation of respective letter signals representing a single letter, to assess the probability of correct word recognition, to terminate the evaluation of respective letter signals and generate second output signals causing said speech output device to output a word obtained with a second desired probability based upon assessment of the probability of correct word recognition, and to carry out speech recognition of the word speech signals using the letter signals as detected and evaluated, if the correct word recognition is not obtained with the second desired probability.

39. (NEW) A communication device for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, comprising:

a data bus;

at least one memory device, coupled to said data bus, to store at least one program;

a speech recognition processor, coupled to said data bus, to detect acoustic word speech signals from a user and to carry out a speech recognition operation;

a speech output device, coupled to said data bus, to produce audio signals simulating speech; and

a central processor, coupled to said data bus, to assess probability of correct speech recognition, to generate first output signals causing said speech output device to prompt the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability, to evaluate letter signals as input by the user, to carry out a word recognition operation following evaluation of respective letter signals representing a single letter based upon assessment of a combined recognition probability using the letter recognition probability and the speech recognition probability, to assess the probability of correct word recognition, and to terminate the evaluation of respective letter signals and generate second output signals causing said speech output device to output a word obtained with a second desired probability based on the assessment of the combined recognition probability.

40. (NEW) The communication device as claimed in claim 39, wherein said central processor also generates a word list based on the combined recognition probability.

41. (NEW) The communication device as claimed in claim 40, wherein said central processor terminates spelling and causes output of the word based solely on a single interrogation as to whether the combined recognition probability is the second desired recognition probability.

42. (NEW) The communication device as claimed in claim 41, wherein upon terminating spelling and outputting the word, said central processor also outputs an appropriate message to the user and terminates detection of the acoustic word speech signals.

43. (NEW) The communication device as claimed in claim 42, wherein, after detection and evaluation of the letter speech signals respectively representing a letter, said central processor also determines whether the user is continuing to speak, and if the user continues to speak the next speech signals respectively representing a letter are detected; while if the user does not continue to speak, said central processor causes outputting of one of the word list and a predetermined number of the words with highest probabilities in the word list.

44. (NEW) The communication device as claimed in claim 43, wherein said communication device is connectable to telephone lines,

further comprising a switching unit coupled to the telephone lines and said data bus.

45. (NEW) An electronically readable data medium storing at least one computer program to control a processor to perform a method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, said method comprising:

detecting acoustic word speech signals from a user;

carrying out a speech recognition operation using a first vocabulary;

assessing probability of correct speech recognition;

prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

detecting and evaluating letter signals as input by the user;

carrying out a word recognition operation, after said detecting and evaluating of respective letter signals representing a single letter, using a second vocabulary larger than the first vocabulary;

assessing the probability of correct word recognition; and

terminating spelling and outputting a word obtained with a second desired probability by said assessing the probability of correct word recognition.

46. (NEW) The electronically readable data medium as claimed in claim 45, wherein the word recognition operation includes

assigning a letter recognition probability based on the letter speech signals; and

determining a word list of all words in the second vocabulary having a letter recognition probability not lower than a highest determined letter recognition probability for any word, minus a first threshold value.

47. (NEW) The electronically readable data medium as claimed in claim 46,

wherein said assessing the probability of correct word recognition comprises determining whether the word list contains only a single word, and

wherein said terminating spelling and outputting the word is performed if only a single word is contained in the word list.

48. (NEW) The electronically readable data medium as claimed in claim 47,

wherein said method further comprises:

carrying out speech recognition of the word speech signals using the word list with each word assigned a speech recognition probability; and

determining whether a highest speech recognition probability and a second highest speech recognition probability differ from one another by a predetermined threshold value; and

wherein if the predetermined threshold value is exceeded by a difference between the highest and second highest speech recognition probabilities, said terminating spelling and outputting the word is performed for the word in the word list with the highest speech recognition probability.

49. (NEW) An electronically readable data medium storing at least one computer program to control a processor to perform a method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, said method comprising:

detecting acoustic word speech signals from a user;

carrying out a speech recognition operation;

assessing probability of correct speech recognition;

prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

detecting and evaluating letter signals as input by the user;

carrying out a word recognition operation, after said detecting and evaluating of respective letter signals representing a single letter;

assessing the probability of correct word recognition;

terminating spelling and outputting a word obtained with a second desired probability by said assessing the probability of correct word recognition; and

carrying out speech recognition of the word speech signals using the letter signals as detected and evaluated, if the correct word recognition is not obtained with the second desired probability.

50. (NEW) An electronically readable data medium storing at least one computer program to control a processor to perform a method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, said method comprising:

detecting acoustic word speech signals from a user;

carrying out a speech recognition operation to obtain a speech recognition probability;

assessing probability of correct speech recognition;

prompting the user to spell out each word for which the probability of correct speech recognition does not reach a first desired probability;

detecting and evaluating letter signals as input by the user of at least one letter, to obtain a letter recognition probability based on each detected letter signal;

carrying out a word recognition operation, after said detecting and evaluating of respective letter signals representing a single letter, based on a combined recognition probability using the letter recognition probability and the speech recognition probability;

assessing the probability of correct word recognition; and

terminating spelling and outputting a word if the word is obtained with a second desired probability by said assessing the probability of correct word recognition.

51. (NEW) The electronically readable data medium as claimed in claim 50, wherein said method further comprises generating a word list based on the combined recognition probability.

52. (NEW) The electronically readable data medium as claimed in claim 51, wherein said terminating spelling and outputting the word is based solely on a single interrogation as to whether the combined recognition probability is the second desired recognition probability.

53. (NEW) The electronically readable data medium as claimed in claim 52, wherein said terminating spelling and outputting the word includes

outputting an appropriate message to the user; and

terminating said detecting the acoustic word speech signals.

54. (NEW) The electronically readable data medium as claimed in claim 53, wherein said method further comprises, after said detecting and evaluating of the letter speech signals respectively representing a letter:

determining whether the user is continuing to speak, and if the user continues to speak the next speech signals respectively representing a letter are detected; and

outputting one of the word list and a predetermined number of the words with highest probabilities in the word list, if the user is not continuing to speak.

IN THE ABSTRACT:

Please DELETE the Abstract in its entirety and replace with the attached Substitute Abstract.

REMARKS

This Preliminary Amendment is submitted to improve the form of the English translation as filed. It is respectfully requested that this Preliminary Amendment be entered in the above-referenced application.

In accordance with the foregoing, claims 1-13 have been canceled and claims 14-54 have been added. Thus, claims 14-54 are pending and are under consideration.

A substitute specification is also being filed herewith. The substitute specification is accompanied by a marked-up copy of the original specification.

If there are any questions regarding these matters, such questions can be addressed by telephone to the undersigned. Otherwise, an early action on the merits is respectfully solicited.

If any further fees are required in connection with the filing of this Preliminary Amendment, please charge same to our Deposit Account No. 19-3935.

Respectfully submitted,

STAAS & HALSEY LLP

Date: 3/1/02

By: Richard A. Gollhofer
Richard A. Gollhofer
Registration No. 31,106

700 Eleventh Street, NW, Suite 500
Washington, D.C. 20001
(202) 434-1500

SUBSTITUTE SPECIFICATION

TITLE OF THE INVENTION

SYSTEM FOR DETECTING AND EVALUATING WORD SPEECH SIGNALS REPRESENTING A WORD FROM A USER OF A SPEECH RECOGNITION SYSTEM

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is based on and hereby claims priority to German Application No. 19942172.2 filed on September 3, 1999 in, the contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] The invention relates to a method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, having the following steps:

- detecting the acoustic word speech signal,
- carrying out a speech recognition operation and assessing the probability of correct speech recognition, and
- if the speech recognition does not reach the desired probability, the user is prompted to spell out the word, and
- detecting and evaluating the letter speech signals spelt out by the user.

[0003] Such a method is known from "Strategies for name recognition in automatic directory assistance systems", Andreas Kellner et al., in 'IEEE Workshop on interactive Voice Technology for Telecommunications Applications (IVTTA), pages 21-26, Turin, Italy, September 1998. A speech recognition system for a telephone network that has a word mode and a spelling mode is described therein. In the word mode, a word is input coherently by being spoken. In the spelling mode, a word is input by spelling out. If a word is not recognized in the word mode with a satisfactory recognition probability, a switchover is made into the spelling mode, in which the word is input by spelling out. Owing to the change into the spelling mode, the word mode can be based on a program that is relatively simple and quick to execute, and it is possible nevertheless to achieve a very high recognition rate, since in the case of all words not recognized exactly, the exact word is input in the spelling mode. However, the price for this high

recognition rate is the user-unfriendly spelling out, which lasts substantially longer than when the corresponding word is uttered coherently.

SUMMARY OF THE INVENTION

[0004] It is an object of the invention to develop the methods of the prior art in a more user friendly fashion.

[0005] The methods according to the invention are distinguished by the following steps:

- carrying out a word recognition operation after the respective detection of the letter speech signals representing a single letter, and
- assessing the probability of correct word recognition, and
- if a word is obtained with the desired probability with the aid of the word recognition, the spelling process is terminated and the word is output.

[0006] In the case of the method according to the invention, an attempt is therefore made to determine the word to be detected as early as after the spelling out of each letter and, if a word is obtained with the desired detection probability, the further spelling process is terminated and the word is output. As a result, the spelling process, which is bothersome for a user, is reduced to a minimum such that the user-friendliness of the method is substantially enhanced by comparison with the known method, and yet an optimal recognition rate is achieved.

[0007] According to a preferred embodiment, a speech recognition operation with the aid of which the word speech signals representing a word from a user, which signals are uttered coherently, is executed on the basis of a smaller vocabulary than in the case of the word recognition operation with the aid of which the letter speech signals representing the individual letters are evaluated. As a result, the computational outlay on the speech recognition operation can be substantially reduced by comparison with a speech recognition operation that takes account of all possible words that may occur. A quick response of the method according to the invention is achieved thereby.

[0008] According to a further preferred embodiment, a renewed speech recognition of the word speech signals is executed in the case of the word recognition operation with the aid of which the letter speech signals are evaluated, the results obtained by the evaluation of the letter speech signals being taken into account in this case. This is performed, for example, by virtue

of the fact that the letter speech signals are used to draw up a word list that is used as vocabulary during the renewed speech recognition.

[0009] The method according to the invention is ended upon termination of the spelling process, and the user of the speech recognition system is output a message to the effect that the spelling process is ended, or the word detected by the word recognition is imparted to him. However, it is also possible for only a predetermined dialog to be continued between the user and the speech recognition system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] These and other objects and advantages of the present invention will become more apparent and more readily appreciated from the following description of the preferred embodiments, taken in conjunction with the accompanying drawings of which:

Fig. 1 is a flowchart of the essential steps of the method according to the invention,

Fig. 2 is a flowchart of the detection and evaluation of the letter speech signals spelt out by the user,

Fig. 3 is a graph indicating how many letters must be spelt on average in order to achieve a predetermined acceptance rate, and

Fig. 4 is a block diagram of a device for executing the method according to the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0011] Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings, wherein like reference numerals refer to like elements throughout.

[0012] The method according to the invention is explained below in more detail with the aid of an exemplary embodiment that is a constituent of an automatic directory inquiry service and which has an automatic speech recognition system for recognizing all German town names.

[0013] 22,077 town names are listed in all German telephone books. These 22,077 words therefore constitute the total vocabulary that contains all town names to be determined with the aid of the speech recognition system.

[0014] Acoustic word speech signals from a user are detected in a step S1 (Fig. 1). The word speech signals are acoustic signals that reproduce a word in a normal, coherent mode of utterance. The words are town names in the present exemplary embodiment.

[0015] The word speech signals detected are evaluated by a speech recognition operation. Such speech recognition operations are known per se. They are used to generate a recognition result that comprises a word or a list of words, the probability of a correct speech recognition being determined in relation to each word and assigned to the respective word.

[0016] A check is made in step S3 as to whether the speech recognition operation was able to determine a word with the desired speech recognition probability. If this is the case, the word determined in step S4, which is a town name in the present exemplary embodiment, is output and the method according to the invention is ended.

[0017] If, by contrast, the result of step S3 is that no word could be determined with the required speech recognition probability, the method sequence goes over to step S5 with the aid of which a spelling process is executed in which the word to be determined is input through spelling out by the user and then evaluated appropriately. The spelling process is explained in more detail below.

[0018] The word determined in step S5 is output in step S6, and the method is ended.

[0019] In the exemplary embodiment relating to detecting and evaluating a town name, the speech recognition is not executed on the basis of the total vocabulary of 22,077 town names, but only on the basis of a medium vocabulary of approximately 1,000 to 5,000 town names. This vocabulary that is considerably reduced by comparison with the total vocabulary includes the town names most frequently asked for. The computer power required to run a computer program for speech recognition can be substantially diminished by the reduction in the vocabulary. The desired town name can be determined very quickly and with a high recognition rate for a majority of the queries because of the reduced vocabulary.

[0020] Since the speech recognition S2 is connected downstream of the spelling process S5, the requirement placed on the probability of a correct speech recognition can be set very high, since a town name incorrectly recognized by the speech recognition can be input again into the spelling process, and a rejection in step S3 does not result in negative effects on the overall result of the method according to the invention.

[0021] The requirements placed on the probability of a correct speech recognition are expediently set so high that the rate of the words that are erroneously recognized in step S2 and are, however, evaluated as correctly recognized in step S3 is smaller than 3% and preferably smaller than 1%.

[0022] The spelling process in accordance with step S5 is illustrated with its individual steps in a flowchart in Fig. 2.

[0023] The user is prompted to fill out the town name in step S7. The user is prompted to spell out the letters individually in the present exemplary embodiment.

[0024] The letter speech signals, which represent an individual letter, are detected and recognized in step S8.

[0025] A word list is drawn up in step S9 in accordance with the letter signals detected and evaluated in step S8. This word list is drawn up on the basis of the total vocabulary of all 22,077 town names, the individual town names being assigned letter recognition probabilities. The letter recognition probability is the probability with which one or more spelt out letters of the word are correctly recognized with the aid of the detected and evaluated letter signals.

[0026] If, for example, the first letter spelt out by the user is a "B", all the town names that begin with a "B" are assigned high letter recognition probabilities. Furthermore, all the town names that start with a "W" are also assigned relatively high letter recognition probabilities, since a "B" and a "W" sound very similar in German, and the "W" can therefore be recognized with a relatively high probability as a correct letter for the detected and evaluated letter signals of the spelt out "B". Town names that begin with another letter are therefore assigned substantially lower letter recognition probabilities. However, only the town names whose letter recognition probability is not lower than the highest determined letter recognition probability minus a predetermined threshold value SW1 are accepted in the word list. The remaining town names are not taken into account in the following steps. The method thereby applied for determining the word list is based on the Viterbi algorithm.

[0027] The list therefore includes all town names that begin with a "B" or a "W" in the case of the input of a "B" as first letter.

[0028] A check is made in the next step S10 as to whether the list includes only a single town name. If this is the case, the method sequence is transferred in step S11 to the main method in

accordance with Fig. 1, where the town name determined is then output in step S6, and the spelling process is terminated.

[0029] If, however, the word list includes a plurality of town names, the method sequence branches to step S12, in which case a speech recognition of the originally input word speech signals is executed anew, the speech recognition being based on the word list drawn up in step S9 as vocabulary. Since the speech recognition is based on the same word speech signals as in step S2, the same speech recognition probabilities are determined for the same words. This step differs from the speech recognition according to step S2 by virtue of the new vocabulary that has been determined from the spelt out letters. The town names newly added by contrast with the vocabulary originally used in step S2 are firstly assigned a speech recognition probability. The speech recognition probability determined by the speech recognition according to step S12 is not combined with the letter recognition probability determined according to the letter recognition of step S8.

[0030] However, it is also possible to combine these two recognition probabilities as an alternative. They can be combined with one another by multiplication, for example.

[0031] A check is made in the next step S13 as to whether the highest speech recognition probability for a town name is higher than the second highest speech recognition probability of a further town name in the word list by a predetermined threshold value SW2. If this is the case, the method sequence branches to step S14 at which the method sequence is transferred again to the main method in which the town name with the highest speech recognition probability is output and the spelling process is terminated. If the highest speech recognition probability is not higher than the next highest speech recognition probability by the predetermined threshold value SW2, the method sequence goes over to step S15 in which the user is output a signal for speaking the next letter, and a pause is made for uttering the next letter. The signal is a short sound signal, for example.

[0032] A check is made in step S16 as to whether the user is speaking a further letter. If the user speaks a further letter, the method sequence goes over again to step S6 at which the further letter is detected and recognized. A loop transversal with the steps S8, S9, S10, S11 or S12, S13, S14 or S15 and S16 is thereby begun.

[0033] As in the first loop transversal, a new word list is drawn up in the case of each further loop transversal in step S9. For this purpose, the individual town names are again assigned a letter recognition probability. This letter recognition probability is determined on the basis of the recognition probabilities with which the individual letters of the town names have been correctly recognized by the detected and evaluated letter signals. The letter recognition probability is calculated by multiplying all the recognition probabilities of the sequence of spelt out letters of the town names for which a corresponding sequence of letter signals has been detected and evaluated. This calculation is executed in such a way that the letter recognition probability previously determined in step S8 is combined with, that is to say multiplied by, the letter recognition probability for the execution of step S8 in the preceding loop transversal.

[0034] Again, included in the word list are only the town names whose letter recognition probability is not lower than the highest determined letter recognition probability minus a predetermined threshold value SW1. The remaining town names are not taken into account in the following steps, that is to say a new list is drawn up, individual town names not being taken into account by comparison with the previous list, and others being included anew. However, a tendency arises in this case in accordance with which the number of the words in the word list is reduced with each loop transversal, since the recognition is the more specific the more letter signals are detected and evaluated.

[0035] In the case of a plurality of transversals of the loop, the word list is reduced solely as determined by the letters input during spelling out, and the interrogation of step S10 is therefore based solely on the letter recognition probability.

[0036] This loop is transversed until, as a result of one of the two interrogations in steps S10 and S13, a town name has been determined with the required recognition probability. If steps S10 and S13 do not lead to termination of the loop, but a termination of the loop ensues in step S16, that is to say because it is established that the user is no longer speaking, the method sequence goes over to step S17 with which either the word that has the highest recognition probability, or a residual list having, for example, the three to ten words on the word list that have been assigned the highest recognition probabilities, is output.

[0037] Illustrated in Fig. 3 is a diagram showing how many letters have to be spelt out on average, that is to say how many loop transversals must be executed until a predetermined acceptance rate is reached. Illustrated by dashes in the diagram is the result for the method

according to the invention, which has two termination criteria at steps S10 and S13. The result for a conventional method without such termination criteria is drawn with a continuous line. It may be seen from this diagram that, for example, after 7 letters have been spelt out an acceptance rate of only just above 40% is achieved with the known spelling methods, whereas an acceptance rate of over 80% is already achieved with the method according to the invention. Substantially fewer letters need be spelt out with the method according to the invention than is the case for conventional spelling methods. This diagram also shows that acceptance rates of 80% to 100% are already achieved with the spelling out of six letters.

[0038] Consequently, the method according to the invention can be used to limit the average number of letters to be spelt out to five to seven.

[0039] A successful recognition rate was already achieved for an average number of 4.9 letters with the aid of the described exemplary embodiment for detecting and evaluating the town names of Germany.

[0040] The present invention is not limited to the automatic detection and recognition of town names, but is suitable, in particular, for all vocabularies with a limited number of words. It can, however, also be used for unlimited vocabularies. The method is then to be modified in a way known per se such that when words that are not yet included in the total vocabulary are being input by spelling them out a routine is executed with the aid of which these words are added to the total vocabulary.

[0041] The method according to the invention can also be modified in such a way that the individual words are assigned a combined recognition probability on the basis of the letter and speech recognition. In this case, the word list is drawn up on the basis of the combined recognition probability. As a consequence of this, the two recognition probabilities of the above described exemplary embodiment, on which the termination criteria according to steps S10 and S13 are based, are identical, for which reason one termination criterion can be deleted.

[0042] According to a simplified method, it is also possible at each loop traversal only to remove words from a word list once drawn up. Since no new words can be added thereby, the speech recognition of step S12 need be executed only the once during the first loop traversal, since town names set forth in the list have all been evaluated already with the corresponding speech recognition probability.

[0043] In the above exemplary embodiment, the letters are uttered in isolation when being spelt out, and are respectively recognized individually. However, it is also possible for the letters to be uttered continuously when being spelt out. In the case of such a continuous method, step S15 (signal for the next letter and pause) can be eliminated.

[0044] Figure 4 shows a device, specifically a telephone communication system 1 for an automatic directory inquiry service. The telephone communication system 1 is designed as a digitally operating telephone communication system having an internal databus 2, a central processor unit 3, a memory unit 4, a speech recognition unit 5 and a speech output unit 6. Connected to the databus 2 is a switching unit 7 via which the analog and digital telephone lines 8, 9 can be switched to connect to the units 3 to 6 via the databus 2. Telephone terminals 10 are connected to the telephone lines 8, 9, it being possible for one or more exchanges 11 to be connected there between.

[0045] A user can select the telephone communication system 1 at one of the telephone terminals 10, by which means he is connected to the units 3 to 6 via the switching unit 7 and the databus 2.

[0046] A plurality of computer programs are stored in the memory unit 4. One of these is provided for executing the above described exemplary embodiment for recognizing town names. If, in his dialog with the telephone communication system, the user reaches a point at which the town name needs to be input, an appropriate prompt is output to the user by the speech output unit 6, and the user then speaks a town name. The latter is then detected with the aid of the speech recognition unit 5 and evaluated in accordance with the above described method, the user being prompted, if required, to spell out the town name.

[0047] Since the method according to the invention is not limited to the recognition of town names, a plurality of programs that operate in accordance with the method according to the invention can be provided, these programs each being capable of recognizing words of specific vocabularies, for example personal names and company names, numbers, stocks or the like. These computer programs designed according to the invention are called up and controlled by a higher level dialog control program.

[0048] These computer programs can also be stored on an electronically readable data medium and, for example, can be transmitted to another telephone communication system.

[0049] The invention has been described in detail with particular reference to preferred embodiments thereof and examples, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention.

ABSTRACT

METHOD FOR DETECTING AND EVALUATING WORD SPEECH SIGNALS REPRESENTING A WORD FROM A USER OF A SPEECH RECOGNITION SYSTEM

In the event of a possibly incorrect speech recognition, the user is prompted to spell out the corresponding word. After each spelt-out word, word recognition is executed such that the spelling process can be terminated given a satisfactory recognition probability.

MARKED-UP COPY OF SUBSTITUTE SPECIFICATION

[Description] TITLE OF THE INVENTION

[METHOD] SYSTEM FOR DETECTING AND EVALUATING WORD SPEECH SIGNALS REPRESENTING A WORD FROM A USER OF A SPEECH RECOGNITION SYSTEM

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is based on and hereby claims priority to German Application No. 19942172.2 filed on September 3, 1999 in, the contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] The invention relates to a method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, having the following steps:

- detecting the acoustic word speech signal,
- carrying out a speech recognition operation and assessing the probability of correct speech recognition, and
- if the speech recognition does not reach the desired probability, the user is prompted to spell out the word, and
- detecting and evaluating the letter speech signals spelt out by the user.

[0003] Such a method is known from "Strategies for name recognition in automatic directory assistance systems", Andreas Kellner et al., in 'IEEE Workshop on interactive Voice Technology for Telecommunications Applications (IVTTA), pages 21-26, Turin, Italy, September 1998. A speech recognition system for a telephone network that has a word mode and a spelling mode is described therein. In the word mode, a word is input coherently by being spoken. In the spelling mode, a word is input by spelling out. If a word is not recognized in the word mode with a satisfactory recognition probability, a switchover is made into the spelling mode, in which the word is input by spelling out. Owing to the change into the spelling mode, the word mode can be based on a program that is relatively simple and quick to execute, and it is possible nevertheless to achieve a very high recognition rate, since in the case of all words not recognized exactly, the exact word is input in the spelling mode. However, the price for this high

recognition rate is the user-unfriendly spelling out, which lasts substantially longer than when the corresponding word is uttered coherently.

SUMMARY OF THE INVENTION

[0004] It is [therefore the] an object of the invention to develop the [method mentioned at the beginning] methods of the prior art in a more user friendly fashion. [The object is achieved by a method having the features of claim 1. Advantageous refinements of the invention are specified in the subclaims.]

[0005] The [method] methods according to the invention [is] are distinguished by the following steps:

- carrying out a word recognition operation after the respective detection of the letter speech signals representing a single letter, and
- assessing the probability of correct word recognition, and
- if a word is obtained with the desired probability with the aid of the word recognition, the spelling process is terminated and the word is output.

[0006] In the case of the [method] methods according to the invention, an attempt is therefore made to determine the word to be detected as early as after the spelling out of each letter and, if a word is obtained with the desired detection probability, the further spelling process is terminated and the word is output. As a result, the spelling process, which is bothersome for a user, is reduced to a minimum such that the user-friendliness of the method is substantially enhanced by comparison with the known method, and yet an optimal recognition rate is achieved.

[0007] According to a preferred embodiment, a speech recognition operation with the aid of which the word speech signals representing a word from a user, which signals are uttered coherently, is executed on the basis of a smaller vocabulary than in the case of the word recognition operation with the aid of which the letter speech signals representing the individual letters are evaluated. As a result, the computational outlay on the speech recognition operation can be substantially reduced by comparison with a speech recognition operation that takes account of all possible words that may occur. A quick response of the method according to the invention is achieved thereby.

[0008] According to a further preferred embodiment, a renewed speech recognition of the word speech signals is executed in the case of the word recognition operation with the aid of which the letter speech signals are evaluated, the results obtained by the evaluation of the letter speech signals being taken into account in this case. This is performed, for example, by virtue of the fact that the letter speech signals are used to draw up a word list that is used as vocabulary during the renewed speech recognition.

[0009] The method according to the invention is ended upon termination of the spelling process, and the user of the speech recognition system is output a message to the effect that the spelling process is ended, or the word detected by the word recognition is imparted to him. However, it is also possible for only a predetermined dialog to be continued between the user and the speech recognition system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] [The] These and other objects and advantages of the present invention [is explained below in more detail with the aid of an exemplary embodiment illustrated in the drawing. In] will become more apparent and more readily appreciated from the following description of the preferred embodiments, taken in conjunction with the accompanying drawings of which:

Fig. 1 [shows] is a flowchart of the essential steps of the method according to the invention, [in a flowchart]

Fig. 2 [shows] is a flowchart of the detection and evaluation of the letter speech signals spelt out by the user, [in a flowchart,]

Fig. 3 [shows a diagram] is a graph indicating how many letters must be spelt on average in order to achieve a predetermined acceptance rate, and

Fig. 4 [shows] is a block diagram of a device for executing the method according to the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0011] Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings, wherein like reference numerals refer to like elements throughout.

[0012] The method according to the invention is explained below in more detail with the aid of an exemplary embodiment that is a constituent of an automatic directory inquiry service and which has an automatic speech recognition system for recognizing all German town names.

[0013] 22,077 town names are listed in all German telephone books. These 22,077 words therefore constitute the total vocabulary that contains all town names to be determined with the aid of the speech recognition system.

[0014] Acoustic word speech signals from a user are detected in a step S1 ([figure] Fig. 1). The word speech signals are acoustic signals that reproduce a word in a normal, coherent mode of utterance. The words are town names in the present exemplary embodiment.

[0015] The word speech signals detected are evaluated by [means of] a speech recognition operation. Such speech recognition operations are known per se. They are used to generate a recognition result that comprises a word or a list of words, the probability of a correct speech recognition being determined in relation to each word and assigned to the respective word.

[0016] A check is made in step S3 as to whether the speech recognition operation was able to determine a word with the desired speech recognition probability. If this is the case, the word determined in step S4, which is a town name in the present exemplary embodiment, is output and the method according to the invention is ended.

[0017] If, by contrast, the result of step S3 is that no word could be determined with the required speech recognition probability, the method sequence goes over to step S5 with the aid of which a spelling process is executed in which the word to be determined is input through spelling out by the user and then evaluated appropriately. The spelling process is explained in more detail below.

[0018] The word determined in step S5 is output in step S6, and the method is ended.

[0019] In the exemplary embodiment relating to detecting and evaluating a town name, the speech recognition is not executed on the basis of the total vocabulary of 22,077 town names, but only on the basis of a medium vocabulary of approximately 1,000 to 5,000 town names. This vocabulary that is considerably reduced by comparison with the total vocabulary includes the town names most frequently asked for. The computer power required to run a computer program for speech recognition can be substantially diminished by the reduction in the vocabulary. The desired town name can be determined very quickly and with a high recognition rate for a majority of the queries because of the reduced vocabulary.

[0020] Since the speech recognition S2 is connected downstream of the spelling process S5, the requirement placed on the probability of a correct speech recognition can be set very high,

since a town name incorrectly recognized by the speech recognition can be input again into the spelling process, and a rejection in step S3 does not result in negative effects on the overall result of the method according to the invention.

[0021] The requirements placed on the probability of a correct speech recognition are expediently set so high that the rate of the words that are erroneously recognized in step S2 and are, however, evaluated as correctly recognized in step S3 is smaller than 3% and preferably smaller than 1%.

[0022] The spelling process in accordance with step S5 is illustrated with its individual steps in a flowchart in [figure] Fig. 2.

[0023] The user is prompted to fill out the town name in step S7. The user is prompted to spell out the letters individually in the present exemplary embodiment.

[0024] The letter speech signals, which represent an individual letter, are detected and recognized in step S8.

[0025] A word list is drawn up in step S9 in accordance with the letter signals detected and evaluated in step S8. This word list is drawn up on the basis of the total vocabulary of all 22,077 town names, the individual town names being assigned letter recognition probabilities. The letter recognition probability is the probability with which one or more spelt out letters of the word are correctly recognized with the aid of the detected and evaluated letter signals.

[0026] If, for example, the first letter spelt out by the user is a "B", all the town names that begin with a "B" are assigned high letter recognition probabilities. Furthermore, all the town names that start with a "W" are also assigned relatively high letter recognition probabilities, since a "B" and a "W" sound very similar in German, and the "W" can therefore be recognized with a relatively high probability as a correct letter for the detected and evaluated letter signals of the spelt out "B". Town names that begin with another letter are therefore assigned substantially lower letter recognition probabilities. However, only the town names whose letter recognition probability is not lower than the highest determined letter recognition probability minus a predetermined threshold value SW1 are accepted in the word list. The remaining town names are not taken into account in the following steps. The method thereby applied for determining the word list is based on the Viterbi algorithm.

[0027] The list therefore includes all town names that begin with a "B" or a "W" in the case of the input of a "B" as first letter.

[0028] A check is made in the next step S10 as to whether the list includes only a single town name. If this is the case, the method sequence is transferred in step S11 to the main method in accordance with [figure] Fig. 1, where the town name determined is then output in step S6, and the spelling process is terminated.

[0029] If, however, the word list includes a plurality of town names, the method sequence branches to step S12, in which case a speech recognition of the originally input word speech signals is executed anew, the speech recognition being based on the word list drawn up in step S9 as vocabulary. Since the speech recognition is based on the same word speech signals as in step S2, the same speech recognition probabilities are determined for the same words. This step differs from the speech recognition according to step S2 by virtue of the new vocabulary that has been determined from the spelt out letters. The town names newly added by contrast with the vocabulary originally used in step S2 are firstly assigned a speech recognition probability. The speech recognition probability determined by the speech recognition according to step S12 is not combined with the letter recognition probability determined according to the letter recognition of step S8.

[0030] However, it is also possible to combine these two recognition probabilities as an alternative. They can be combined with one another by [means of] multiplication, for example.

[0031] A check is made in the next step S13 as to whether the highest speech recognition probability for a town name is higher than the second highest speech recognition probability of a further town name in the word list by a predetermined threshold value SW2. If this is the case, the method sequence branches to step S14 at which the method sequence is transferred again to the main method in which the town name with the highest speech recognition probability is output and the spelling process is terminated. If the highest speech recognition probability is not higher than the next highest speech recognition probability by the predetermined threshold value SW2, the method sequence goes over to step S15 in which the user is output a signal for speaking the next letter, and a pause is made for uttering the next letter. The signal is a short sound signal, for example.

[0032] A check is made in step S16 as to whether the user is speaking a further letter. If the user speaks a further letter, the method sequence goes over again to step S6 at which the further letter is detected and recognized. A loop transversal with the steps S8, S9, S10, S11 or S12, S13, S14 or S15 and S16 is thereby begun.

[0033] As in the first loop transversal, a new word list is drawn up in the case of each further loop transversal in step S9. For this purpose, the individual town names are again assigned a letter recognition probability. This letter recognition probability is determined on the basis of the recognition probabilities with which the individual letters of the town names have been correctly recognized by the detected and evaluated letter signals. The letter recognition probability is calculated by multiplying all the recognition probabilities of the sequence of spelt out letters of the town names for which a corresponding sequence of letter signals has been detected and evaluated. This calculation is executed in such a way that the letter recognition probability previously determined in step S8 is combined with, that is to say multiplied by, the letter recognition probability for the execution of step S8 in the preceding loop transversal.

[0034] Again, included in the word list are only the town names whose letter recognition probability is not lower than the highest determined letter recognition probability minus a predetermined threshold value SW1. The remaining town names are not taken into account in the following steps, that is to say a new list is drawn up, individual town names not being taken into account by comparison with the previous list, and others being included anew. However, a tendency arises in this case in accordance with which the number of the words in the word list is reduced with each loop transversal, since the recognition is the more specific the more letter signals are detected and evaluated.

[0035] In the case of a plurality of transversals of the loop, the word list is reduced solely as determined by the letters input during spelling out, and the interrogation of step S10 is therefore based solely on the letter recognition probability.

[0036] This loop is transversed until, as a result of one of the two interrogations in steps S10 and S13, a town name has been determined with the required recognition probability. If steps S10 and S13 do not lead to termination of the loop, but a termination of the loop ensues in step S16, that is to say because it is established that the user is no longer speaking, the method sequence goes over to step S17 with which either the word that has the highest recognition

probability, or a residual list having, for example, the three to ten words on the word list that have been assigned the highest recognition probabilities, is output.

[0037] Illustrated in [figure] Fig. 3 is a diagram showing how many letters have to be spelt out on average, that is to say how many loop transversals must be executed until a predetermined acceptance rate is reached. Illustrated by dashes in the diagram is the result for the method according to the invention, which has two termination criteria at steps S10 and S13. The result for a conventional method without such termination criteria is drawn with a continuous line. It may be seen from this diagram that, for example, after 7 letters have been spelt out an acceptance rate of only just above 40% is achieved with the known spelling methods, whereas an acceptance rate of over 80% is already achieved with the method according to the invention. Substantially fewer letters need be spelt out with the method according to the invention than is the case for conventional spelling methods. This diagram also shows that acceptance rates of 80% to 100% are already achieved with the spelling out of six letters.

[0038] Consequently, the method according to the invention can be used to limit the average number of letters to be spelt out to five to seven.

[0039] A successful recognition rate was already achieved for an average number of 4.9 letters with the aid of the described exemplary embodiment for detecting and evaluating the town names of Germany.

[0040] The present invention is not limited to the automatic detection and recognition of town names, but is suitable, in particular, for all vocabularies with a limited number of words. It can, however, also be used for unlimited vocabularies. The method is then to be modified in a way known per se such that when words that are not yet included in the total vocabulary are being input by spelling them out a routine is executed with the aid of which these words are added to the total vocabulary.

[0041] The method according to the invention can also be modified in such a way that the individual words are assigned a combined recognition probability on the basis of the letter and speech recognition. In this case, the word list is drawn up on the basis of the combined recognition probability. As a consequence of this, the two recognition probabilities of the above described exemplary embodiment, on which the termination criteria according to steps S10 and S13 are based, are identical, for which reason one termination criterion can be deleted.

[0042] According to a simplified method, it is also possible at each loop traversal only to remove words from a word list once drawn up. Since no new words can be added thereby, the speech recognition of step S12 need be executed only the once during the first loop traversal, since town names set forth in the list have all been evaluated already with the corresponding speech recognition probability.

[0043] In the above exemplary embodiment, the letters are uttered in isolation when being spelt out, and are respectively recognized individually. However, it is also possible for the letters to be uttered continuously when being spelt out. In the case of such a continuous method, step S15 (signal for the next letter and pause) can be eliminated.

[0044] Figure 4 shows a device, specifically a telephone communication system 1 for an automatic directory inquiry service. The telephone communication system 1 is designed as a digitally operating telephone communication system having an internal databus 2, a central processor unit 3, a memory unit 4, a speech recognition unit 5 and a speech output unit 6. Connected to the databus 2 is a switching unit 7 via which the analog and digital telephone lines 8, 9 can be switched to connect to the units 3 to 6 via the databus 2. Telephone terminals 10 are connected to the telephone lines 8, 9, it being possible for one or more exchanges 11 to be connected there between.

[0045] A user can select the telephone communication system 1 at one of the telephone terminals 10, by which means he is connected to the units 3 to 6 via the switching unit 7 and the databus 2.

[0046] A plurality of computer programs are stored in the memory unit 4. One of these is provided for executing the above described exemplary embodiment for recognizing town names. If, in his dialog with the telephone communication system, the user reaches a point at which the town name needs to be input, an appropriate prompt is output to the user by [means of] the speech output unit 6, and the user then speaks a town name. The latter is then detected with the aid of the speech recognition unit 5 and evaluated in accordance with the above described method, the user being prompted, if required, to spell out the town name.

[0047] Since the method according to the invention is not limited to the recognition of town names, a plurality of programs that operate in accordance with the method according to the invention can be provided, these programs each being capable of recognizing words of specific vocabularies, for example personal names and company names, numbers, stocks or the like.

These computer programs designed according to the invention are called up and controlled by a higher level dialog control program.

[0048] These computer programs can also be stored on an electronically readable data medium and, for example, can be transmitted to another telephone communication system.

[0049] The invention has been described in detail with particular reference to preferred embodiments thereof and examples, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention.

4/prts

GR 99 P 2738

Description

Method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system

The invention relates to a method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, having the following steps:

- detecting the acoustic word speech signal,
- carrying out a speech recognition operation and assessing the probability of correct speech recognition, and
- if the speech recognition does not reach the desired probability, the user is prompted to spell out the word, and
- detecting and evaluating the letter speech signals spelt out by the user.

Such a method is known from "Strategies for name recognition in automatic directory assistance systems", Andreas Kellner et al., in 'IEEE Workshop on interactive Voice Technology for Telecommunications Applications (IVTTA), pages 21-26, Turin, Italy, September 1998. A speech recognition system for a telephone network that has a word mode and a spelling mode is described therein. In the word mode, a word is input coherently by being spoken. In the spelling mode, a word is input by spelling out. If a word is not recognized in the word mode with a satisfactory recognition probability, a switchover is made into the spelling mode, in which the word is input by spelling out. Owing to the change into the spelling mode, the word mode can be based on a program that is relatively simple and quick to execute, and it is possible nevertheless to achieve a very high recognition rate, since in the case of all words not recognized exactly, the exact word

is input in the spelling mode. However, the price for this high recognition rate is the user-unfriendly spelling out, which lasts substantially longer than when the corresponding word is uttered coherently.

US 5,638,425 discloses a method for detecting and evaluating word speech signals representing a word from a user of an automatic call information service. The method has a word mode and a spelling mode. If, in the word mode, the speech recognition does not recognize an acoustic speech signal with a desired probability, the method switches over into the spelling mode and prompts the user to spell out the word, for which purpose the user spells out the entire word or parts thereof.

Kaspar B. et al., "Spracherkennung für großes Vokabular durch Buchstabieren" ["Speech recognition for a large vocabulary by means of spelling"], ITG Fachberichte, April 28th, 1986, pages 31-36 discloses an automatic call information service having a spelling mode that carries out a word recognition operation after the respective detection of letter speech signals representing a letter, in which case the spelling process is terminated and the word is output if a word with the desired recognition probability is obtained by the word recognition.

It is the object of the invention to develop said methods in a user-friendly fashion.

The object is achieved by means of methods, a device and a computer program product having the features of the independent claims. Advantageous refinements of the invention are specified in the subclaims.

The methods include the following steps:

- carrying out a word recognition operation after the respective detection of the letter speech signals representing a single letter, and

- assessing the probability of correct word recognition, and
- if a word is obtained with the desired probability with the aid of the word recognition, the spelling process is terminated and the word is output.

In the case of the methods, an attempt is therefore made to determine the word to be detected as early as after the spelling out of each letter and, if a word is obtained with the desired detection probability, the further spelling process is terminated and the word is output. As a result, the spelling process, which is bothersome for a user, is reduced to a minimum such that the user-friendliness of the method is substantially enhanced by comparison with the known method, and yet an optimal recognition rate is achieved.

Moreover, a speech recognition operation with the aid of which the word speech signals representing a word from a user, which signals are uttered coherently, is executed on the basis of a smaller vocabulary than in the case of the word recognition operation with the aid of which the letter speech signals representing the individual letters are evaluated. As a result, the computational outlay on the speech recognition operation can be substantially reduced by comparison with a speech recognition operation that takes account of all possible words that may occur. A quick response of the method according to the invention is achieved thereby.

Alternatively or in addition, a renewed speech recognition of the word speech signals is executed in the case of the word recognition operation with the aid of which the letter speech signals are evaluated, the results obtained by the evaluation of the letter speech signals being taken into account in this case. This is performed, for example, by virtue of the fact that the letter speech signals are used to draw up a word list that is used as vocabulary during the renewed speech recognition.

The method according to the invention is ended upon termination of the spelling process, and the user of the speech recognition system is output a message to the effect that the spelling process is ended, or the word detected by the word recognition is imparted to him. However, it is also possible for only a predetermined dialog to be continued between the user and the speech recognition system.

The invention is explained below in more detail with the aid of an exemplary embodiment illustrated in the drawing. In the drawings:

Fig. 1 shows the essential steps of the method according to the invention, in a flowchart,

Fig. 2 shows the detection and evaluation of the letter speech signals spelt out by the user, in a flowchart,

Fig. 3 shows a diagram indicating how many letters must be spelt on average in order to achieve a predetermined acceptance rate, and

Fig. 4 shows a device for executing the method according to the invention.

The method according to the invention is explained below in more detail with the aid of an exemplary embodiment that is a constituent of an automatic directory inquiry service and which has an automatic speech recognition system for recognizing all German town names.

22,077 town names are listed in all German telephone books. These 22,077 words therefore constitute the total vocabulary that contains all town names to be determined with the aid of the speech recognition system.

Acoustic word speech signals from a user are detected in a step S1 (figure 1). The word speech signals are acoustic signals that reproduce a word in a normal, coherent mode of utterance. The words are town names in the present exemplary embodiment.

The word speech signals detected are evaluated by means of a speech recognition operation. Such speech recognition operations are known per se. They are used to generate a recognition result that comprises a word or a list of words, the probability of a correct speech recognition being determined in relation to each word and assigned to the respective word.

A check is made in step S3 as to whether the speech recognition operation was able to determine a word with the desired speech recognition probability. If this is the case, the word determined in step S4, which is a town name

in the present exemplary embodiment, is output and the method according to the invention is ended.

If, by contrast, the result of step S3 is that no word could be determined with the required speech recognition probability, the method sequence goes over to step S5 with the aid of which a spelling process is executed in which the word to be determined is input through spelling out by the user and then evaluated appropriately. The spelling process is explained in more detail below.

The word determined in step S5 is output in step S6, and the method is ended.

In the exemplary embodiment relating to detecting and evaluating a town name, the speech recognition is not executed on the basis of the total vocabulary of 22,077 town names, but only on the basis of a medium vocabulary of approximately 1,000 to 5,000 town names. This vocabulary that is considerably reduced by comparison with the total vocabulary includes the town names most frequently asked for. The computer power required to run a computer program for speech recognition can be substantially diminished by the reduction in the vocabulary. The desired town name can be determined very quickly and with a high recognition rate for a majority of the queries because of the reduced vocabulary.

Since the speech recognition S2 is connected downstream of the spelling process S5, the requirement placed on the probability of a correct speech recognition can be set very high, since a town name incorrectly recognized by the speech recognition can be input again into the spelling process, and a rejection in step S3 does not result in negative effects on the overall result of the method according to the invention.

The requirements placed on the probability of a correct speech recognition are expediently set so high that the rate of the words that are erroneously recognized in step S2 and are, however, evaluated as correctly recognized in step S3 is smaller than 3% and preferably smaller than 1%.

The spelling process in accordance with step S5 is illustrated with its individual steps in a flowchart in figure 2.

The user is prompted to fill out the town name in step S7. The user is prompted to spell out the letters individually in the present exemplary embodiment.

The letter speech signals, which represent an individual letter, are detected and recognized in step S8.

A word list is drawn up in step S9 in accordance with the letter signals detected and evaluated in step S8. This word list is drawn up on the basis of the total vocabulary of all 22,077 town names, the individual town names being assigned letter recognition probabilities. The letter recognition probability is the probability with which one or more spelt out letters of the word are correctly recognized with the aid of the detected and evaluated letter signals.

If, for example, the first letter spelt out by the user is a "B", all the town names that begin with a "B" are assigned high letter recognition probabilities. Furthermore, all the town names that start with a "W" are also assigned relatively high letter recognition probabilities, since a "B" and a "W" sound very similar in German, and the "W" can therefore be recognized with a relatively high probability as a correct letter for the detected and evaluated letter signals of the spelt out "B". Town names that begin with another letter are therefore assigned substantially lower letter recognition probabilities. However, only the town names whose letter recognition probability is not lower than the highest determined letter recognition probability minus a predetermined threshold value SW1 are accepted in the word list. The remaining town names are not taken into account in the following steps. The method thereby applied for determining the word list is based on the Viterbi algorithm.

The list therefore includes all town names that begin with a "B" or a "W" in the case of the input of a "B" as first letter.

A check is made in the next step S10 as to whether the list

includes only a single town name. If this is the case, the method sequence is transferred in step S11 to the main method in accordance with figure 1, where the town name determined is then output in step S6, and the spelling process is terminated.

If, however, the word list includes a plurality of town names, the method sequence branches to step S12, in which case a speech recognition of the originally input word speech signals is executed anew, the speech recognition being based on the word list drawn up in step S9 as vocabulary. Since the speech recognition is based on the same word speech signals as in step S2, the same speech recognition probabilities are determined for the same words. This step differs from the speech recognition according to step S2 by virtue of the new vocabulary that has been determined from the spelt out letters. The town names newly added by contrast with the vocabulary originally used in step S2 are firstly assigned a speech recognition probability. The speech recognition probability determined by the speech recognition according to step S12 is not combined with the letter recognition probability determined according to the letter recognition of step S8.

However, it is also possible to combine these two recognition probabilities as an alternative. They can be combined with one another by means of multiplication, for example.

A check is made in the next step S13 as to whether the highest speech recognition probability for a town name is higher than the second highest speech recognition probability of a further town name in the word list by a predetermined threshold value SW2. If this is the case, the method sequence branches to step S14 at which the method sequence is transferred again to the main method in which the town name with the highest speech recognition probability is output and the spelling process is terminated. If the highest speech recognition probability is not higher than the next highest speech recognition probability by the predetermined threshold value SW2, the method sequence goes over to step S15 in which the user is output a signal for speaking the next letter, and a pause is made for uttering the next letter. The signal is a

short sound signal, for example.

A check is made in step S16 as to whether the user is speaking a further letter. If the user speaks a further letter, the method sequence goes over again to step S6 at which the further letter is detected and recognized. A loop transversal with the steps S8, S9, S10, S11 or S12, S13, S14 or S15 and S16 is thereby begun.

As in the first loop transversal, a new word list is drawn up in the case of each further loop transversal in step S9. For this purpose, the individual town names are again assigned a letter recognition probability. This

letter recognition probability is determined on the basis of the recognition probabilities with which the individual letters of the town names have been correctly recognized by the detected and evaluated letter signals. The letter recognition probability is calculated by multiplying all the recognition probabilities of the sequence of spelt out letters of the town names for which a corresponding sequence of letter signals has been detected and evaluated. This calculation is executed in such a way that the letter recognition probability previously determined in step S8 is combined with, that is to say multiplied by, the letter recognition probability for the execution of step S8 in the preceding loop transversal.

Again, included in the word list are only the town names whose letter recognition probability is not lower than the highest determined letter recognition probability minus a predetermined threshold value SW1. The remaining town names are not taken into account in the following steps, that is to say a new list is drawn up, individual town names not being taken into account by comparison with the previous list, and others being included anew. However, a tendency arises in this case in accordance with which the number of the words in the word list is reduced with each loop transversal, since the recognition is the more specific the more letter signals are detected and evaluated.

In the case of a plurality of transversals of the loop, the word list is reduced solely as determined by the letters input during spelling out, and the interrogation of step S10 is therefore based solely on the letter recognition probability.

This loop is transversed until, as a result of one of the two interrogations in steps S10 and S13, a town

name has been determined with the required recognition probability. If steps S10 and S13 do not lead to termination of the loop, but a termination of the loop ensues in step S16, that is to say because it is established that the user is no longer speaking, the method sequence goes over to step S17 with which either the word that has the highest recognition probability, or a residual list having, for example, the three to ten words on the word list that have been assigned the highest recognition probabilities, is output.

Illustrated in figure 3 is a diagram showing how many letters have to be spelt out on average, that is to say how many loop transversals must be executed until a predetermined acceptance rate is reached. Illustrated by dashes in the diagram is the result for the method according to the invention, which has two termination criteria at steps S10 and S13. The result for a conventional method without such termination criteria is drawn with a continuous line. It may be seen from this diagram that, for example, after 7 letters have been spelt out an acceptance rate of only just above 40% is achieved with the known spelling methods, whereas an acceptance rate of over 80% is already achieved with the method according to the invention. Substantially fewer letters need be spelt out with the method according to the invention than is the case for conventional spelling methods. This diagram also shows that acceptance rates of 80% to 100% are already achieved with the spelling out of six letters.

Consequently, the method according to the invention can be used to limit the average number of letters to be spelt out to five to seven.

A successful recognition rate was already achieved for an average number of 4.9 letters

with the aid of the described exemplary embodiment for detecting and evaluating the town names of Germany.

The present invention is not limited to the automatic detection and recognition of town names, but is suitable, in particular, for all vocabularies with a limited number of words. It can, however, also be used for unlimited vocabularies. The method is then to be modified in a way known per se such that when words that are not yet included in the total vocabulary are being input by spelling them out a routine is executed with the aid of which these words are added to the total vocabulary.

The method according to the invention can also be modified in such a way that the individual words are assigned a combined recognition probability on the basis of the letter and speech recognition. In this case, the word list is drawn up on the basis of the combined recognition probability. As a consequence of this, the two recognition probabilities of the above described exemplary embodiment, on which the termination criteria according to steps S10 and S13 are based, are identical, for which reason one termination criterion can be deleted.

According to a simplified method, it is also possible at each loop traversal only to remove words from a word list once drawn up. Since no new words can be added thereby, the speech recognition of step S12 need be executed only the once during the first loop traversal, since town names set forth in the list have all been evaluated already with the corresponding speech recognition probability.

In the above exemplary embodiment, the letters are uttered in isolation when being spelt out, and are respectively recognized individually. However, it is

also possible for the letters to be uttered continuously when being spelt out. In the case of such a

continuous method, step S15 (signal for the next letter and pause) can be eliminated.

Figure 4 shows a device, specifically a telephone communication system 1 for an automatic directory inquiry service. The telephone communication system 1 is designed as a digitally operating telephone communication system having an internal databus 2, a central processor unit 3, a memory unit 4, a speech recognition unit 5 and a speech output unit 6. Connected to the databus 2 is a switching unit 7 via which the analog and digital telephone lines 8, 9 can be switched to connect to the units 3 to 6 via the databus 2. Telephone terminals 10 are connected to the telephone lines 8, 9, it being possible for one or more exchanges 11 to be connected there between.

A user can select the telephone communication system 1 at one of the telephone terminals 10, by which means he is connected to the units 3 to 6 via the switching unit 7 and the databus 2.

A plurality of computer programs are stored in the memory unit 4. One of these is provided for executing the above described exemplary embodiment for recognizing town names. If, in his dialog with the telephone communication system, the user reaches a point at which the town name needs to be input, an appropriate prompt is output to the user by means of the speech output unit 6, and the user then speaks a town name. The latter is then detected with the aid of the speech recognition unit 5 and evaluated in accordance with the above described method, the user being prompted, if required, to spell out the town name.

Since the method according to the invention is not limited to the recognition of town names, a plurality

of programs that operate in accordance with the method according to the invention can be provided, these programs each being capable of recognizing words of specific

vocabularies, for example personal names and company names, numbers, stocks or the like. These computer programs designed according to the invention are called up and controlled by a higher level dialog control program.

5

These computer programs can also be stored on an electronically readable data medium and, for example, can be transmitted to another telephone communication system.

Patent claims

1. A method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, having the following steps:

- detecting the acoustic word speech signals,

- carrying out a speech recognition operation and assessing the probability of correct speech recognition,

- if the speech recognition does not reach the desired probability, the user is prompted to spell out the word,

- detecting and evaluating the letter signals spelt out by the user,

- carrying out a word recognition operation after the respective detection of the letter signals representing a single letter,

- assessing the probability of correct word recognition,

- if a word is obtained with the desired probability with the aid of the word recognition, the spelling process is terminated and the word is output,

- the speech recognition operation being executed on the basis of a smaller vocabulary than does the word recognition operation.

2. The method as claimed in claim 1, characterized in that the during the word recognition operation, a word list is drawn up in accordance with the detected letter speech signals, the words of a total vocabulary in each case being assigned a letter recognition probability on the basis of the letter speech signals, and the word list comprising all words whose letter recognition probability is not lower than the highest determined letter recognition probability of a word minus a threshold value (SW1).

3. The method as claimed in claim 2, characterized in that a check is made as to whether the word list contains only a single word, and if only a single word is contained the latter is output and the spelling process is terminated.

4. The method as claimed in claim 2 or 3, characterized in that a renewed speech recognition of the word speech signals is carried out in which the words of the word list are respectively assigned a speech recognition probability, and a check is made as to whether the highest and the second highest speech recognition probability differ from one another by a predetermined threshold value (SW2), and if this is the case the word of the word list with the highest speech recognition probability is output, and the spelling process is terminated.

5. A method for detecting and evaluating word speech signals representing a word from the user of a speech recognition system, in particular as claimed in one of the preceding claims, having the following steps:

- detecting the acoustic word speech signals,
- carrying out a speech recognition operation and assessing the probability of correct speech recognition,
- if the speech recognition does not reach the desired probability, the user is prompted to spell out the word,
- detecting and evaluating the letter signals spelt out by the user,
- carrying out a word recognition operation after the respective detection of the letter signals representing a single letter,

-assessing the probability of correct word recognition,

- if a word is obtained with the desired probability with the aid of the word recognition, the spelling process is terminated and the word is output,

and

- a renewed speech recognition of the word speech signals is carried out taking account of the detected letter speech signals.

6. A method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system, in particular as claimed in one of the preceding claims, having the following steps:

- detecting the acoustic word speech signals,

- carrying out a speech recognition operation and assessing the probability of correct speech recognition,

- if the speech recognition does not reach the desired probability, the user is prompted to spell out the word,

- detecting and evaluating the letter signals spelt out by the user,

- carrying out a word recognition operation after the respective detection of the letter signals representing a single letter,

- assessing the probability of correct word recognition,

- if a word is obtained with the desired probability with the aid of the word recognition, the spelling process is terminated and the word is output,

- during the word recognition operation a letter recognition probability being determined on the basis of the detected and evaluated letter signals and being combined with a speech recognition probability determined on the basis of the detected and

evaluated word speech signals to form a combined recognition probability.

7. The method as claimed in claim 6, characterized in that a word list is drawn up in accordance with the combined recognition probability.

8. The method as claimed in claim 6 or 7, characterized in that a check is made solely with the aid of a single interrogation as to whether a word is obtained with the desired recognition probability, the combined recognition probability being used as recognition probability.

9. The method as claimed in one of claims 1 to 8, characterized in that the spelling process is terminated by outputting an appropriate message to the user and by ending the method for detecting and evaluating a word.

10. The method as claimed in one of claims 2 to 9, characterized in that when the spelling process has not yet been terminated, after the detection and evaluation of the letter speech signals respectively representing a letter, a check is made as to whether the user is continuing to speak, and if he is continuing to speak the next speech signals respectively representing a letter are detected, and if the user is not continuing to speak the word list or a predetermined number of the words with the highest probability of the word list is output.

11. A device that is set up, and has means, for carrying out a method as claimed in one of claims 1 to 10.

12. The device as claimed in claim 11, characterized in that

the device is a telephone communication system (1) that has a switching unit (7) with the aid of which telephone lines (8, 9) can be connected to the internal databus (2).

13. A computer program product for a data processing system that contains software code sections with the aid of which a method as claimed in at least one of claims 1 to 10 can be executed on a data processing system.

GR 99 P 2738

Abstract

Method for detecting and evaluating word speech signals representing a word from a user of a speech recognition system

In the event of a possibly incorrect speech recognition, the method according to the invention prompts the user to spell out the corresponding word. After each spelt-out word, word recognition is executed such that the spelling process can be terminated given a satisfactory recognition probability.

Figure 2

FIG 1

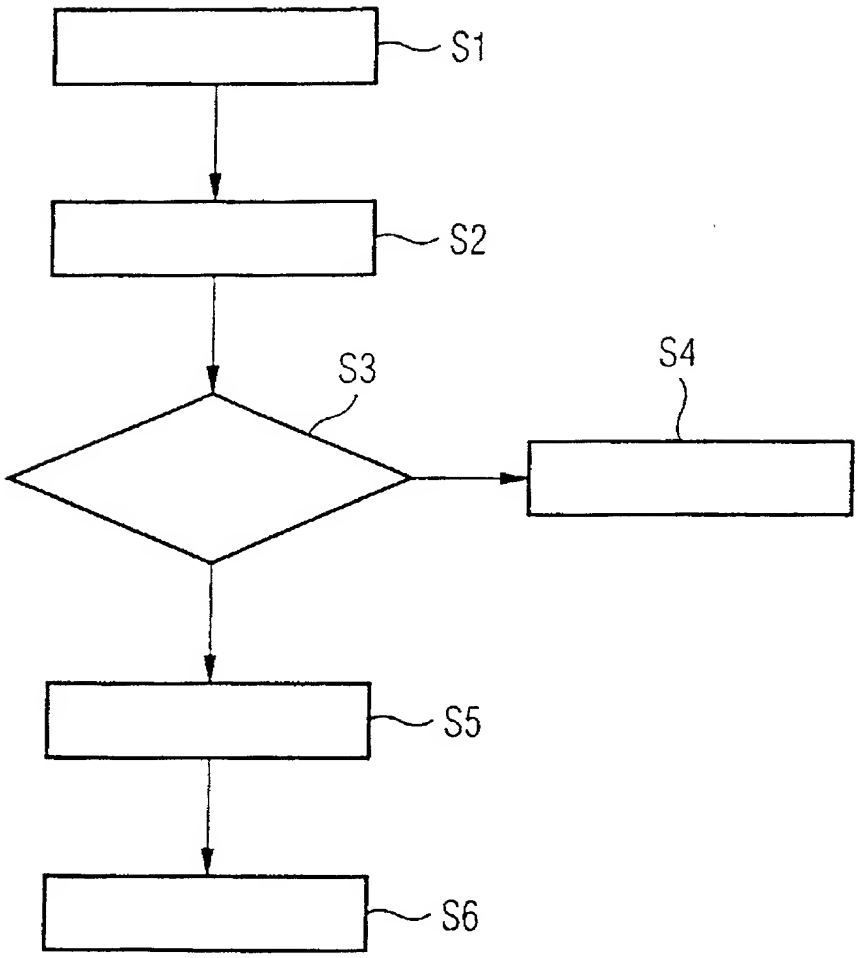
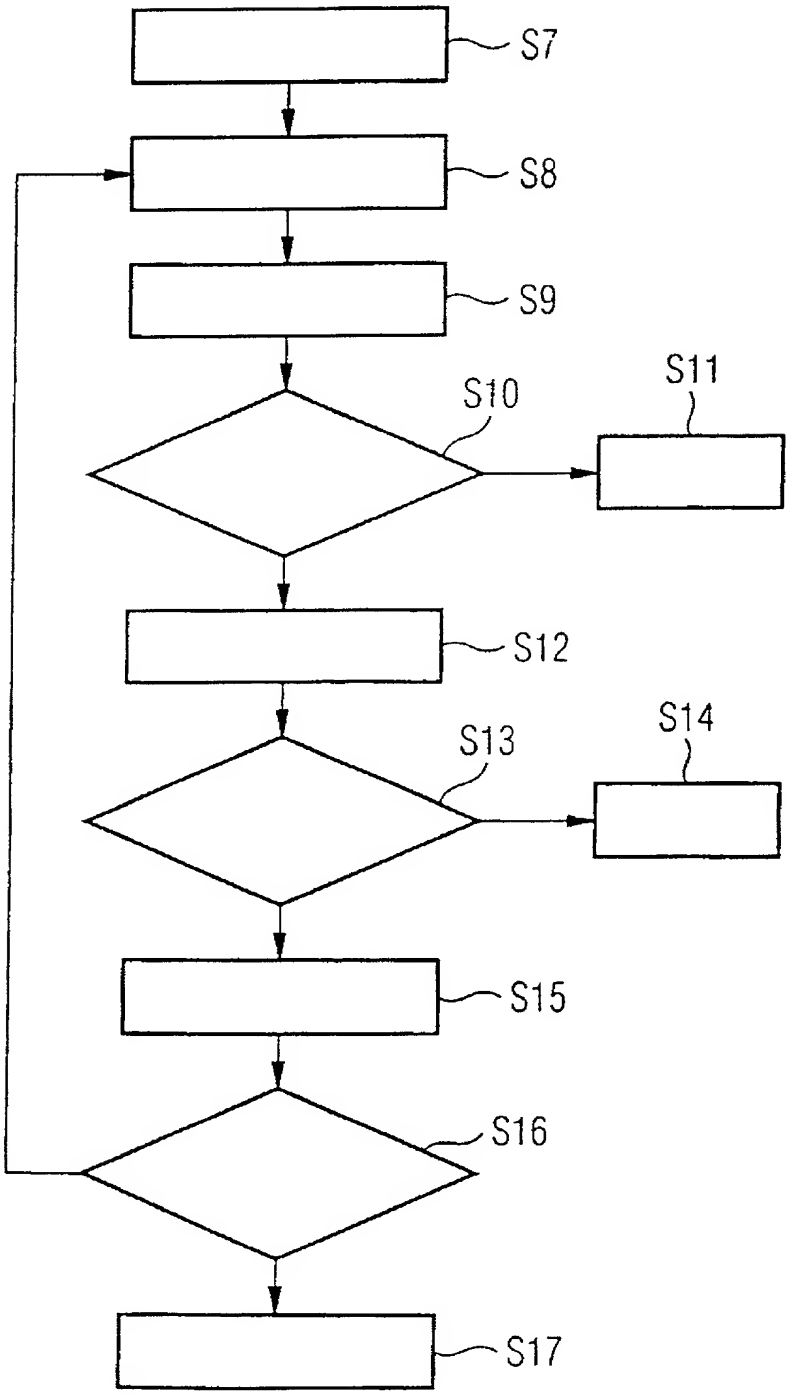


FIG 2



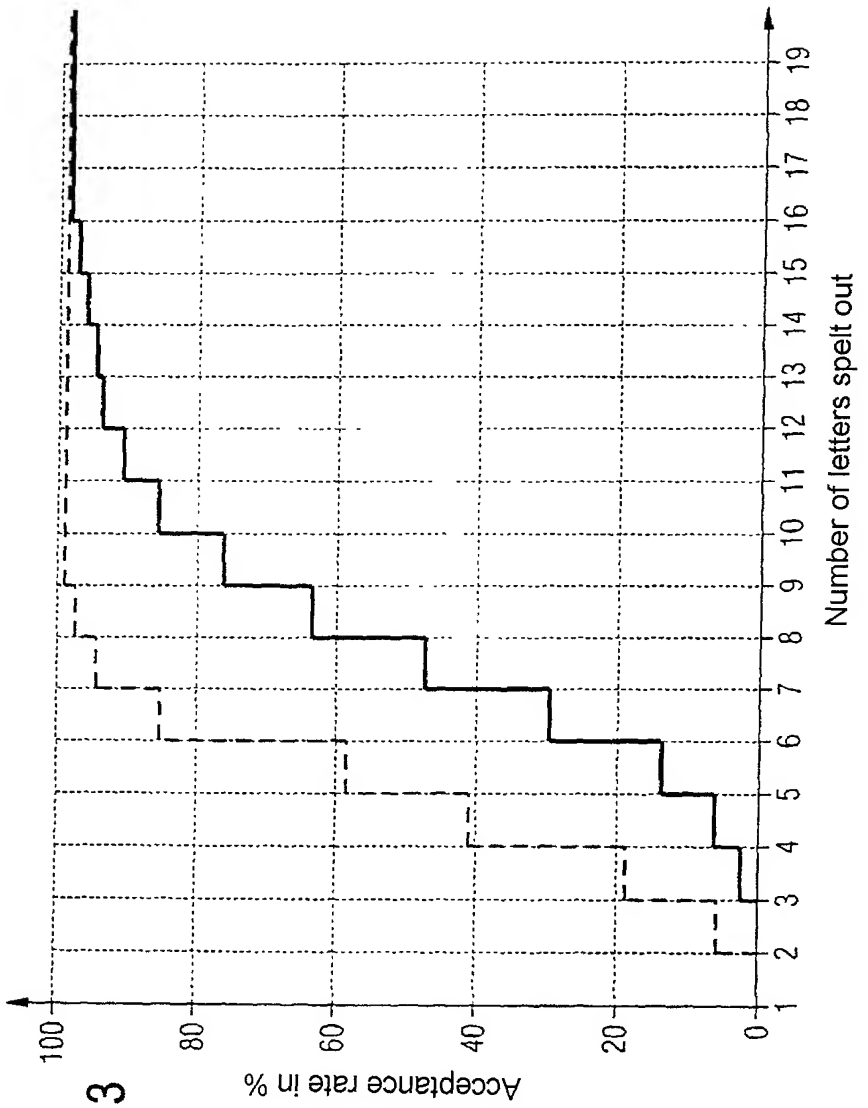
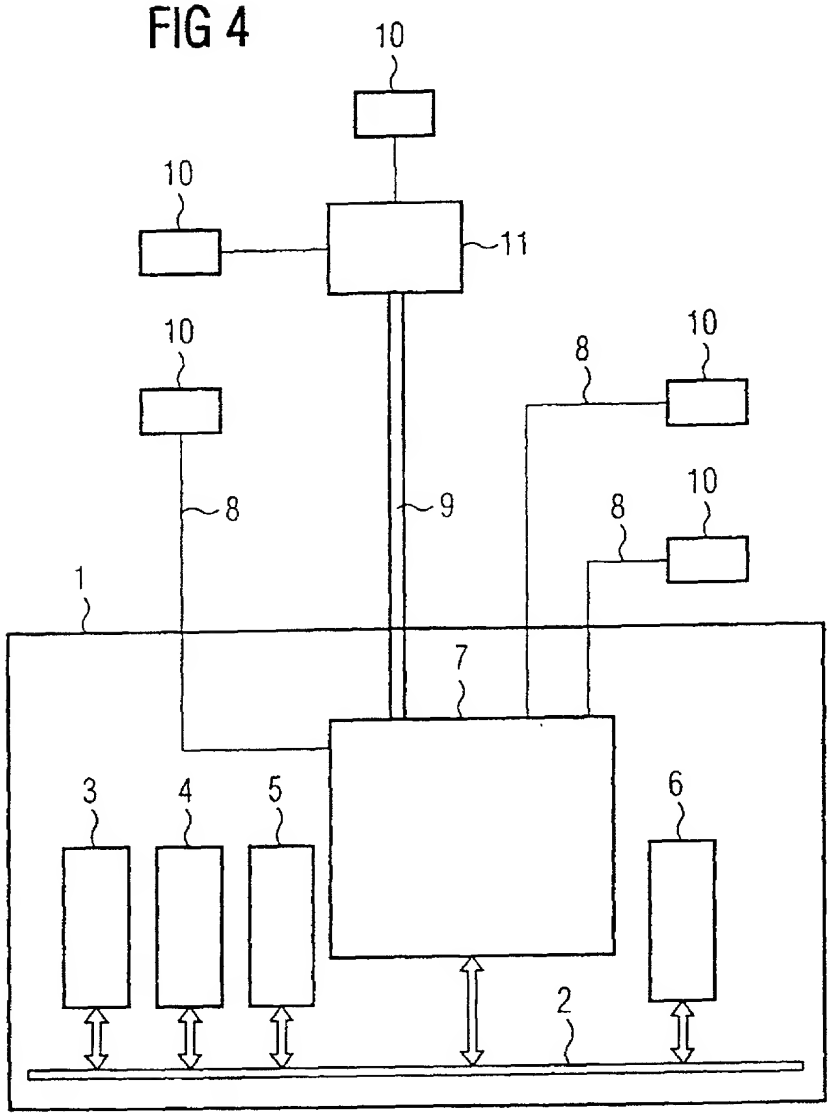


FIG 3

FIG 4



Declaration and Power of Attorney For Patent Application
Erklärung Für Patentanmeldungen Mit Vollmacht
German Language Declaration

Als nachstehend benannter Erfinder erkläre ich hiermit an Eides Statt:

As a below named inventor, I hereby declare that:

dass mein Wohnsitz, meine Postanschrift, und meine Staatsangehörigkeit den im Nachstehenden nach meinem Namen aufgeführten Angaben entsprechen,

My residence, post office address and citizenship are as stated below next to my name,

dass ich, nach bestem Wissen der ursprüngliche, erste und alleinige Erfinder (falls nachstehend nur ein Name angegeben ist) oder ein ursprünglicher, erster und Miterfinder (falls nachstehend mehrere Namen aufgeführt sind) des Gegenstandes bin, für den dieser Antrag gestellt wird und für den ein Patent beantragt wird für die Erfindung mit dem Titel:

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

Verfahren zum Erfassen und Auswerten von ein Wort darstellenden Wortsprachsignalen eines Benutzers eines Spracherkennungssystems

Method and device for detecting and evaluating vocal signals representing a word emitted by a user of a voice-recognition system

deren Beschreibung

(zutreffendes ankreuzen)
☐ hier beigefügt ist.
☒ am 31.05.2000 als
PCT internationale Anmeldung
PCT Anmeldungsnummer PCT/DE00/01787
eingereicht wurde und am _____
abgeändert wurde (falls tatsächlich abgeändert).

the specification of which

(check one)
☐ is attached hereto.
☒ was filed on 31.05.2000 as
PCT international application
PCT Application No. PCT/DE00/01787
and was amended on _____
(if applicable)

Ich bestätige hiermit, dass ich den Inhalt der obigen Patentanmeldung einschliesslich der Ansprüche durchgesehen und verstanden habe, die eventuell durch einen Zusatzantrag wie oben erwähnt abgeändert wurde.

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims as amended by any amendment referred to above.

Ich erkenne meine Pflicht zur Offenbarung irgendwelcher Informationen, die für die Prüfung der vorliegenden Anmeldung in Einklang mit Absatz 37, Bundesgesetzbuch, Paragraph 1.56(a) von Wichtigkeit sind, an.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

Ich beanspruche hiermit ausländische Prioritätsvorteile gemäss Abschnitt 35 der Zivilprozessordnung der Vereinigten Staaten, Paragraph 119 aller unten angegebenen Auslandsanmeldungen für ein Patent oder eine Erfindersurkunde, und habe auch alle Auslandsanmeldungen für ein Patent oder eine Erfindersurkunde nachstehend gekennzeichnet, die ein Anmeldedatum haben, das vor dem Anmeldedatum der Anmeldung liegt, für die Priorität beansprucht wird.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

German Language Declaration

Prior foreign appplications
Priorität beansprucht

			Priority Claimed	
19942172.2	DE	03.09.1999	☒	☐
(Number) (Nummer)	(Country) (Land)	(Day Month Year Filed) (Tag Monat Jahr eingereicht)	Yes Ja	No Nein
			☐	☐
(Number) (Nummer)	(Country) (Land)	(Day Month Year Filed) (Tag Monat Jahr eingereicht)	Yes Ja	No Nein
			☐	☐
(Number) (Nummer)	(Country) (Land)	(Day Month Year Filed) (Tag Monat Jahr eingereicht)	Yes Ja	No Nein

Ich beanspruche hiermit gemäss Absatz 35 der Zivilprozessordnung der Vereinigten Staaten, Paragraph 120, den Vorzug aller unten aufgeführten Anmeldungen und falls der Gegenstand aus jedem Anspruch dieser Anmeldung nicht in einer früheren amerikanischen Patentanmeldung laut dem ersten Paragraphen des Absatzes 35 der Zivilprozeßordnung der Vereinigten Staaten, Paragraph 122 offenbart ist, erkenne ich gemäss Absatz 37, Bundesgesetzbuch, Paragraph 1.56(a) meine Pflicht zur Offenbarung von Informationen an, die zwischen dem Anmeldedatum der früheren Anmeldung und dem nationalen oder PCT internationalen Anmeldedatum dieser Anmeldung bekannt geworden sind.

I hereby claim the benefit under Title 35. United States Code. §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §122, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occured between the filing date of the prior application and the national or PCT international filing date of this application.

PCT/DE00/01787 (Application Serial No.) (Anmeldeseriennummer)	31.05.2000 (Filing Date D, M, Y) (Anmeldedatum T, M, J)	anhängig (Status) (patentiert, anhängig, aufgegeben)	pending (Status) (patented, pending abandoned)
_____ (Application Serial No.) (Anmeldeseriennummer)	(Filing Date D,M,Y) (Anmeldedatum T, M; J)	_____ (Status) (patentiert, anhängig, aufgeben)	(Status) (patented, pending, abandoned)

Ich erkläre hiermit, dass alle von mir in der vorliegenden Erklärung gemachten Angaben nach meinem besten Wissen und Gewissen der vollen Wahrheit entsprechen, und dass ich diese eidesstattliche Erklärung in Kenntnis dessen abgebe, dass wissentlich und vorsätzlich falsche Angaben gemäss Paragraph 1001, Absatz 18 der Zivilprozessordnung der Vereinigten Staaten von Amerika mit Geldstrafe belegt und/oder Gefängnis bestraft werden koennen, und dass derartig wissentlich und vorsätzlich falsche Angaben die Gültigkeit der vorliegenden Patentanmeldung oder eines darauf erteilten Patentes gefährden können.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

German Language Declaration

VERTRETUNGSVOLLMACHT: Als benannter Erfinder beauftrage ich hiermit den nachstehend benannten Patentanwalt (oder die nachstehend benannten Patentanwälte) und/oder Patent-Agenten mit der Verfolgung der vorliegenden Patentanmeldung sowie mit der Abwicklung aller damit verbundenen Geschäfte vor dem Patent- und Warenzeichenamt: *(Name und Registrationsnummer anführen)*

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. *(list name and registration number)*

And I hereby appoint

Customer No. 21171

Telefongespräche bitte richten an: *(Name und Telefonnummer)*

Direct Telephone Calls to: *(name and telephone number)*

Ext. _____

Postanschrift:

Send Correspondence to:

Staas & Halsey LLP
700 Eleventh Street NW, Suite 500 20001 Washington, DC
Telephone: (001) 202 434 1500 and Facsimile (001) 202 434 1501

or

Customer No. 21171

Voller Name des einzigen oder ursprünglichen Erfinders:	Full name of sole or first inventor:
JOSEF BAUER	JOSEF BAUER
Unterschrift des Erfinders Datum: Josef Bauer 23.1.2002	Inventor's signature Date
Wohnsitz: MUENCHEN, DEUTSCHLAND	Residence: MUENCHEN, GERMANY DEX
Staatsangehörigkeit: DE	Citizenship: DE
Postanschrift: WINZERER STR. 96 80797 MUENCHEN	Post Office Addess: WINZERER STR. 96 80797 MUENCHEN
Voller Name des zweiten Miterfinders (falls zutreffend): Dr. JOCHEN JUNKAWITSCH	Full name of second joint inventor, if any: Dr. JOCHEN JUNKAWITSCH
Unterschrift des Erfinders Datum: Jochen Junkawitsch 5 Feb. 2002	Second Inventor's signature Date
Wohnsitz: Eschenbach, DEUTSCHLAND	Residence: Eschenbach, GERMANY DEX
Staatsangehörigkeit: DE	Citizenship: DE
Postanschrift: Sonnenstr. 24 92676 Eschenbach	Post Office Address: Sonnenstr. 24 92676 Eschenbach

(Bitte entsprechende Informationen und Unterschriften im Falle von dritten und weiteren Miterfindern angeben).

(Supply similar information and signature for third and subsequent joint inventors).

Voller Name des dritten Miterfinders:	Full name of third joint inventor:
Dr. TOBIAS SCHNEIDER	Dr. TOBIAS SCHNEIDER
Unterschrift des Erfinders Datum T. Schneider 14.02.02	Inventor's signature Date
Wohnsitz MUENCHEN, DEUTSCHLAND	Residence MUENCHEN, GERMANY DEX
Staatsangehörigkeit DE	Citizenship DE
Postanschrift KRANZHORNSTR. 7 81825 MUENCHEN	Post Office Address KRANZHORNSTR. 7 81825 MUENCHEN
Voller Name des vierten Miterfinders:	Full name of fourth joint inventor:
Unterschrift des Erfinders Datum	Inventor's signature Date
Wohnsitz ,	Residence ,
Staatsangehörigkeit	Citizenship
Postanschrift	Post Office Address
Voller Name des fünften Miterfinders:	Full name of fifth joint inventor:
Unterschrift des Erfinders Datum	Inventor's signature Date
Wohnsitz ,	Residence ,
Staatsangehörigkeit	Citizenship
Postanschrift	Post Office Address
Voller Name des sechsten Miterfinders:	Full name of sixth joint inventor:
Unterschrift des Erfinders Datum	Inventor's signature Date
Wohnsitz ,	Residence ,
Staatsangehörigkeit	Citizenship
Postanschrift	Post Office Address

(Bitte entsprechende Informationen und Unterschriften im Falle von dritten und weiteren Miterfindern angeben).

(Supply similar information and signature for third and subsequent joint inventors).